The Gap Closure Plan must be presented by the CEO of SBC/Ameritech to the CEO of the affected CLEC's, in addition to a designated representative of the Illinois Commerce Commission staff.

The purpose of this requirement is purely to gain exposure by SBC/Ameritech Executive management as opposed to checks simply being cut.

7. Application to Remedy Tiers

a. **Parity Measure Remedies For Tier I**- Remedies for parity measurements are based upon statistical comparison of service performance levels provided to each CLEC, compared to service levels provided by SBC/Ameritech to retail customers and to SBC/Ameritech's affiliate. T he CLECs believe that the intent of the Telecommunications Act of 1996 is clear – SBC/Ameritech must provide parity service to CLECs as compared to its treatment of affiliates as well as its retail customers. Therefore, the CLECs propose that remedies would be due for parity measures that show either superior retail or affiliate treatment compared to wholesale performance. Performance levels are based upon evaluation of the modified z-score statistic (z) as defined in the Local Competition Users Service Group document "Statistical Tests for Local Service Parity." (See attached). The modified z-score is a statistic that is calculated from retail and wholesale performance data that can be used as an index to test whether retail and wholesale performance are substantially the same. If the modified z score is less than a critical value, as determined below, then the statistical test signals that a disparity of service exists between wholesale and retail performance. The CLECs propose that for all sufficiently disaggregated submeasures that the critical value be determined in a manner that balances the probability of Type I (ILEC found guilty when innocent) and Type II (ILEC found innocent when guilty) error probabilities. Since a fixed critical value will not accomplish this, the CLECs have agreed to use the balancing methodology proposed in Appendix C of the Statistician's joint filing for Louisiana Public Service Commission (LPSC) Docket U-22252 Subdocket C[2], hereto attached as Attachment 1, to detect discrimination in all

[2] Statistical Techniques For The Analysis And Comparison Of Performance Measurement Data. Submitted to Louisiana Public Service Commission (LPSC) Docket U-22252 Subdocket C.

submeasures. Since the appendix performs the calculation for the more general case of a truncated z score and deeply disaggregated submeasures, we have also attached a specific calculation for use with the modified z score as defined in this plan for use in the State of Illinois with its set of performance measures (See Attachment 2.) The CLECs propose the use of this methodology with a delta value of 0.25.[3]Incorporating this delta along with the number of data points collected by submeasure, a balancing critical value, z*, is easily calculated for each remediable submeasure. When the modified z-score statistic is compared to the balancing critical value, a sample size independent test occurs which automatically balances the Type I and Type II error probabilities. Furthermore, the ratio of the modified z-score to the balancing critical value is an explicitly sample size independent measure of the severity of the miss, which is used to escalate remedy dollar amounts in this proposal.

Furthermore, in order to increase computational stability and avoid potential gaming, the CLECs propose that remedy amounts should be a continuous function of severity, once disparity is declared by the test. In the CLEC proposal a simple quadratic function is used to easily compute these dollar amounts.

MAGNITUDE PAYMENTS FOR PARITY MEASURE MISSES

Range of modified z-statistic value (z)	Performance Designation	Applicable Consequence ($)
greater than or equal z*	Compliant	0
less than z* to 5z*/3	Basic Failure	

[3] Delta is a standardized measure of material difference between ILEC performance for its retail or affiliate compared to the ILECs whole performance for the CLECs. . The 0.25 delta chosen is a compromise position. Some CLECs were concerned that 0.25 was too generous and that CLECs could still be harmed competitively without remedy using this delta. The CLECs agreed to the joint proposal as an opportunity to study the impact of the 0.25 delta pending the six month review of the plan.

less than $5z^*/3$ to $3z^*$	Intermediate Failure	$a(z/z^*)^2 + b(z/z^*) + c$
less than $3z^*$	Severe Failure	25,000[4]

$a = 5625$

$b = -11250$

$c = 8125.$

In this table it is assumed that a submeasure is worse as its value gets larger and that the definition of modified z score (z) is the same as in the Texas business rules.

b. **Benchmark Measure Remedy for Tier I** - Remedies for benchmark measures are based upon a comparison of achieved service performance levels for CLECs to the established benchmarks. The benchmark levels were established at the lower end of acceptable performance in order to provide the minimum acceptable level of service that would allow the CLECs to compete. These levels should therefore be met 100% of the time. However, to account for random variation, engineering compromises, etc., the benchmark proportions (B) are set at less than 100% depending on the submeasure. Therefore, the resulting benchmark proportions should be considered a "bright line" limit that SBC/Ameritech must meet, and no further statistical considerations are needed. Although further statistical considerations would lead to multiple mitigation in a remedy plan, it would be unfair to order the ILEC to satisfy the benchmarks when sample sizes are small. In such cases a small sample size table is included for benchmarks in this proposal.

Service performance levels that do not achieve the benchmarks are subject to remedy payments. The CLECs have compromised on the values in the charts below. The dollar amounts take into account that the remedies associated with missing a strict benchmark proportion (e.g., 99%) should escalate faster than remedies associated with a less strict benchmark proportion (e.g., 90%).

[4] The levels in the plan will need to be revisited as market entry increases, particularly with the availability of UNE-P and EELs products. At some point, these per measure remedy levels will become an inadequate deterrent to discrimination when CLEC ordering volumes are high.

MAGNITUDE PAYMENTS FOR BENCHMARK MEASURE MISSES

CLEC Data Set Size	Benchmark Percentage Adjustments for Small Data Sets (Applicable to Data Sets < 30)		
	85.0%	**90.0%**	**95.0%**
5	80.0%	80.0%	80.0%
6	83.3%	83.3%	83.3%
7	85.0%	85.7%	85.7%
8	75.0%	87.5%	87.5%
9	77.8%	88.9%	88.9%
10	80.0%	90.0%	90.0%
20	85.0%	90.0%	95.0%
30	83.3%	90.0%	93.3%

Range of Benchmark Result (x)	Performance Designation	Applicable Consequence ($)
Meets or exceeds B%	Compliant	0
Meets or exceeds (1.5B-50)% but worse than B%	Basic Failure	$d[x/(100-B)]^2 + eB[x/(100-B)^2] + f[B/(100-B)]^2 + g$
Meets or exceeds (2B-100)% but worse than (1.5B-50)%	Intermediate Failure	
Worse than (2B-100)%	Severe Failure	25,000

Where,

d = 22500

e = -45000

f = 22500

g = 2500

c. **Parity Measure Remedies for Tier II** - The same rules apply under Tier II to the aggregate (or pooled) data of the individual CLECs as are employed for the individual CLEC data under Tier I, except that a more lenient 5z*/3 critical value is used.

Range of modified z-statistic value (z)	Performance Designation	Applicable Consequence ($)
greater than or equal 5z*/3	Indeterminate	0
less than 5z*/3 to 3z*	Market Impacting	$n\,[a(z/z^*)^2 + b(z/z^*) + c]$
less than 3z*	Market Constraining	n25,000

The value for "n" should be determined based upon the most recent data for the state and relating to resold lines and UNE loops as reported in the Report of Local Competition published by the FCC. The calculation would be based on the most current data reported to the FCC and be as follows: (resold lines + UNE loops)/(total switched lines). This will give the percentage of SBC/Ameritech Illinois switched lines purchased by CLECs. The result represents the level of competition in the state of Illinois

Lines provided to CLECs/Total SBC/Ameritech and CLEC Lines	Value of "n"
more than 50%	0
more than 40% less than or equal 50%	1
more than 30% less than or equal 40%	2
more than 20% less than or equal 30%	4
more than 10% less than or equal 20%	6
more than 5% less than or equal 10%	8
0% to less than or equal 5%	10

Thus, as competition becomes established, the size of the applicable Tier II consequence is reduced to zero if SBC/Ameritech no longer provides a majority of the local lines to the CLECs in its serving area. Based upon current data, the current value of "n" for SBC/Ameritech Illinois is 10.

d. **Benchmark Measure Remedies for Tier II -** The same rules apply under Tier II to the aggregate (or pooled) data of the individual CLECs as are employed for the individual

CLEC data under Tier I, except that consequences do not apply until the pooled CLEC performance results degrades to a point that is equivalent to an intermediate failure designation.

TIER II REMEDY PAYMENTS - BENCHMARK MEASURES

Range of Benchmark Result (x)	Failure Designation	Applicable Consequence ($)
Meets or exceeds (1.5B-50)%	Indeterminate	0
Meets or exceeds (2B-100)% but worse than (1.5B-50)%	Market Impacting	$n\,\{d[x/(100\text{-}B)]^2 + eB[x/(100\text{-}B)^2] + f[B/(100\text{-}B)]^2 + g\}$
Worse than (2B-100)%	Market Constraining	n25,000

e. **Chronic Remedy Payments** - Regardless of the type of measurement (parity or benchmark), if performance fails to achieve the Compliant level in consecutive reporting periods, then additional consequences should apply. The recommended treatment for chronic failures is to assess a chronic failure over-ride in the third consecutive month of non-compliant performance. When the chronic failure override applies, a consequence equal to a 'Severe Failure" ($25,000 per chronic failure per month) for Tier I and "Market Constraining" (n$25,000 per chronic failure per month) for Tier II should apply until such time as performance for the specific measurement result is again classified as Compliant.

f. **Review Threshold** – In addition to establishing an overall review threshold at 36% net local return, regulatory review also would be triggered without withholding remedies in escrow for any month where SBC/Ameritech's remedy payments exceed 1/6 of $125M, or $ 20.8M. The review would focus on discovering the source of SBC/Ameritech's poor performance, and on how the Commission could incent compliant performance promptly, which may include additional remedies or other consequences such as a recommendation that the FCC suspend or not grant 271 relief and/or marketing.

g. **Reporting** -- Remedies are applicable to non-regulatory approved late reports, incomplete reports (missing sub-metrics) and late corrective action reports where they are applicable. These payments will be made to the State of Illinois. These remedies are outlined below.

h. **<u>Late Reports</u>**

Late Reports Per Day	$5,000
Incomplete Reports Per Submetric Per Day	$1,000
Late Corrective Action Reports	$5,000
Late Or Missing Change Management Notices for Metrics and/or Unauthorized* Noticed Changes	$5,000

*Unauthorized means change made unilaterally by SBC/Ameritech without agreement from CLEC collaborative participants.

i. **Reporting Structure:**

SBC/Ameritech Illinois retail data shall be compared to individual CLEC data and, separately, to aggregate CLEC data that excludes the affiliate data. Additionally, SBC/Ameritech's affiliate data shall be compared to individual CLEC data and, separately, to aggregate CLEC data.

CLECs shall have the right to review SBC/Ameritech data, and SBC/Ameritech affiliate raw data, subject to an appropriate protective agreement.

III. Application and Payment of Performance Remedies

A. The remedy plan supplements remedies already included in CLEC interconnection agreements. CLECs also may voluntarily negotiate additions, deletions or changes to the metrics adopted in this collaborative for inclusion in interconnection agreements. Upon issuance of an order by the Commission approving this remedy plan, the metrics developed and remedies would be in force for all CLECs buying service through tariff or interconnection agreement from SBC/Ameritech. A CLEC wishing to be subject to the remedy plan would be required to notify SBC/Ameritech and the Commission in writing and the CLEC's "opt -in" would become effective 20 days from the date of said written notice. Voluntarily negotiated amendments to the remedy plan must also be filed with the Commission and would be automatically approved unless rejected by the Commission within 30 days of filing.

B. Performance remedy payments will be determined on a monthly basis and will be applied at a submeasure level for each CLEC for each failed submeasure.

C. Performance measures and remedies apply to all types of CLEC services, regardless of mode of entry, including but not limited to special access and high capacity services.

D. Payments to the CLECs will be made by check by the end of the month following the data report (e.g. June data, reported in July, remedies paid by August 31). An invoice will accompany the payment explaining the calculation of each submetric missed (base and any magnitude or duration remedies should be specified). Payment by check is necessary in order to ensure certain payment and is easier for the CLECs to administer and track. Bill credits are inappropriate as they are not easily traceable to a specific CLEC account for credit, are less visible to SBC/Ameritech executives and hence less likely to incent improvement and are hard to track when SBC/Ameritech billing is erratic or subject to numerous billing disputes. Remedies for prior periods also can potentially be greater than the bill for a given month. It is counterintuitive to require CLECs to buy additional services from a vendor to receive full compensation for past inferior performance.

E. Participation in this remedy plan does not affect a CLEC's right to bring a separate action before a state commission, the Federal Communications Commission, or the courts for a violation by SBC-Ameritech of the Telecommunications Act of 1996. The existence of this plan similarly does not affect a state commission's authority under either federal or state law to hear such an action or commence such an action on its own initiative, and to redress such a violation in the form of damages or official findings.

F. To the extent the same performance measures are reported on a regional basis by SBC/Ameritech and any of the State PUCs or FCC makes a finding that SBC/Ameritech misreports wholesale data, the Commission may fine SBC/Ameritech up to $10,000 per misreported performance measure.

IV. Mitigation Measures and Dispute Resolution

The use of statistical testing employing the balancing methodology provides a reasonable level of deviation from a strict parity requirement and helps equalize the effects of random variation among all parties. For parity measures that represent worse performance when they have larger values, a Tier 1 modified z score less than 0 indicates that the CLEC received poorer average performance than SBC/Ameritech provided for itself within the monthly sampled data. Therefore, if we declare disparity when the value of the modified z score, as calculated from the data, is below the (negative) balancing value (z^*) we provide the only mitigation required. For Tier 2 performance measures, which have still more negative critical value ($5z^*/3$) of the modified z test for the aggregated CLEC data, mitigation is even greater. However, remedies are potentially greater on declaration of disparity. No additional mitigation (such as a k-table) is required, which greatly simplifies the operation, directness, and understandability of the plan.

SBC/Ameritech will perform a limited root-cause analysis process at a CLEC's request for chronic performance failures.

Either SBC/Ameritech or the CLEC may initiate a request for an expedited hearing process to resolve differences associated with performance parity and remedy payment issues; however, payments must continue to the CLECs pending the outcome of such proceeding.

V. Audits

A. Annual Audit

SBC/Ameritech will support (i.e., pay for) an annual compre hensive audit of its reporting procedures and reportable data. SBC/Ameritech will include all systems, processes and procedures associated with the production and reporting of performance measurement results. A third party auditor will perform this audit . SBC/Ameritech and the CLECs will jointly select the third party auditor. If the parties cannot agree on the auditor, the auditors selected by each party will jointly determine the auditor. Costs for these annual audits will be borne by SBC/Ameritech.

The comprehensive Annual Audits will be conducted every twelve (12) months, with the first such audit commencing twelve (12) months after the conclusion of the KPMG OSS Test's metric replication. (At its completion, SBC/Ameritech shall submit its annual c omprehensive audit to the Commission and distribute copies to CLECs.

B. Mini – Audits:

In addition to an annual audit, the CLECs would have the right to mini -audits of individual performance measures/submeasures during the year. When a CLEC has reason t o believe the data collected for a measure is flawed or the reporting criteria for the measure is not being adhered to, it has the right to have a mini -audit performed on the specific measure/sub-measure upon written request (including e-mail), which will include the designation of a CLEC representative to engage in discussions with SBC/Ameritech about the requested mini-audit. If, 30 days after the CLEC's written request, the CLEC believes that the issue has not been resolved to its satisfaction, the CLEC will commence the mini-audit upon providing SBC/Ameritech with 5 business days advance written notice. Each CLEC would be limited to auditing three single measures/sub-measures or one domain area (preorder, ordering, provisioning, maintenance or billing) during the audit year. The audit year shall commence with the start of the KPMG OSS test (or an Annual Audit). Mini -Audits may be requested for months including and subsequent to the month in which the KPMG OSS or an Annual Audit was initiated. Mini -audits cannot be requested by a CLEC while the OSS third party test or an Annual Audit is being conducted (i.e. before completion).

Mini-Audits will include all systems, processes and procedures associated with the production and reporting of performance measurement results for the audited measure/sub-measure. Mini-Audits will include two (2) months of data, and all parties agree that raw data supporting the performance measurement results will be available monthly to CLECs.

No more than three (3) Mini-Audits will be conducted simultaneously unless more than one CLEC wants the same measure/sub-measure audited at the same time, in which case, Mini-Audits of the same measure/sub-measure shall count as one Mini-Audit for the purposes of this paragraph only.

A third party auditor, selected by the same method as described above, will conduct mini-Audits. SBC/Ameritech will pay for fifty percent (50%) of the costs of the mini-audits. The other fifty percent (50%) of the costs will be divided among the CLEC(s) requesting the mini-audit unless SBC/Ameritech is found to be "materially" misreporting or misrepresenting data or to have non-compliant procedures, in which case, SBC/Ameritech would pay for the entire cost of the third party auditor. SBC/Ameritech will be deemed "materially" at fault when a reported successful measure changes as a consequence of the audit to a missed measure, or there is a change from an ordinary missed measure to intermediate or severe. Each party to the Mini-Audit shall bear its own internal costs, regardless of which party ultimately bears the costs of the third party auditor.

If, during a Mini-Audit, it is found that for more than 30% of the measures in a major service category SBC/Ameritech is "materially" at fault (i.e., a reported successful measure changes as a consequence of the audit to a missed measure, or there is a change from an ordinary missed measure to intermediate or severe), the entire service category will be re-audited at the expense of SBC/Ameritech. The major service categories for this purpose are:

- Pre-Ordering/Ordering
- Billing
- Provisioning - POTS and UNE Loop and Port Combinations
- Provisioning - Resale Specials and UNE Loop and Port Combinations
- Provisioning - Unbundled Network Elements
- Maintenance - POTS and UNE Loop and Port Combinations
- Maintenance - Resale Specials and UNE Loop and Port Combinations
- Maintenance - Unbundled Network Elements
- Interconnection Trunks
- Local Number Portability
- Database - 911
- Database - Directory Assistance
- Database - NXX
- Collocation
- Coordinated Conversions

Each Mini-Audit shall be submitted to the CLEC involved and to the Commission as a proprietary document. SBC/Ameritech will provide notification to all CLECs of any Mini-Audit requested when the request for the audit is made.

Appendix C
Balancing the Type I and Type II Error Probabilities
of the
Truncated Z Test Statistic

This appendix describes the methodology for balancing the error probabilities when the Truncated Z statistic, described in Appendix A, is used for performance measure parity testing. There are four key elements of the statistical testing process:

1. the null hypothesis, H_0, that parity exists between ILEC and CLEC services
2. the alternative hypothesis, H_a, that the ILEC is giving better service to its own customers
3. the Truncated Z test statistic, Z^T, and
4. a critical value, c

The decision rule[5] is

- If $Z^T < c$ then accept H_a.
- If $Z^T \geq c$ then accept H_0.

There are two types of error possible when using such a decision rule:

Type I Error: Deciding favoritism exists when there is, in fact, no favoritism.
Type II Error: Deciding parity exists when there is, in fact, favoritism.

The probabilities of each type of each are:

Type I Error: $\alpha = P(Z^T < c \mid H_0)$.

Type II Error: $\beta = P(Z^T \geq c \mid H_a)$.

In what follows, we show how to find a balancing critical value, c_B, so that $\alpha = \beta$.

General Methodology

The general form of the test statistic that is being used is

$$z_0 = \frac{\hat{T} - E(\hat{T}|H_0)}{SE(\hat{T}|H_0)}, \tag{C.1}$$

where

$\hat{T}$ is an estimator that is (approximately) normally distributed,

$E(\hat{T} \mid H_0)$ is the expected value (mean) of $\hat{T}$ under the null hypothesis, and

[5] This decision rule assumes that the smaller a performance measure is, the better the service. If the opposite is true, then reverse the decision rule.

$\mathrm{SE}(\hat{T} \mid H_0)$ is the standard error of $\hat{T}$ under the null hypothesis.

Thus, under the null hypothesis, z_0 follows a standard normal distribution. However, this is not true under the alternative hypothesis. In this case,

$$z_a = \frac{\hat{T} - E(\hat{T}|H_a)}{\mathrm{SE}(\hat{T}|H_a)}$$

has a standard normal distributio n. Here

$E(\hat{T} \mid H_a)$ is the expected value (mean) of $\hat{T}$ under the alternative hypothesis, and

$\mathrm{SE}(\hat{T} \mid H_a)$ is the standard error of $\hat{T}$ under the alternative hypothesis.

Notice that

$$\begin{aligned}\beta &= P(z_0 > c \mid H_a) \\ &= P\left(z_a > \frac{c\,\mathrm{SE}(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{\mathrm{SE}(\hat{T} \mid H_a)} \right)\end{aligned} \tag{C.2}$$

and recall that for a standard normal random variable z and a constant b, $P(z < b) = P(z > -b)$. Thus,

$$\alpha = P(z_0 < c) = P(z_0 > -c) \tag{C.3}$$

Since we want $\alpha = \beta$, the right hand sides of (C.2) and (C.3) represent the same area under the standard normal density. Therefore, it must be the case that

$$-c = \frac{c\,\mathrm{SE}(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{\mathrm{SE}(\hat{T} \mid H_a)}.$$

Solving this for c give the general formula for a balancing critical value:

$$c_B = \frac{E(\hat{T} \mid H_a) - E(\hat{T} \mid H_0)}{\mathrm{SE}(\hat{T} \mid H_a) + \mathrm{SE}(\hat{T} \mid H_0)} \tag{C.4}$$

The Balancing Critical Value of the Truncated Z

In Appendix A, the Truncated Z statistic is defined as

$$Z^T = \frac{\sum_j W_j Z_j^* - \sum_j W_j E(Z_j^* \mid H_0)}{\sqrt{\sum_j W_j^2 \mathrm{Var}(Z_j^* \mid H_0)}}$$

In terms of equation (C.1) we have

$$\hat{T} = \sum_j W_j Z_j^*$$
$$E(\hat{T} \mid H_0) = \sum_j W_j E(Z_j^* \mid H_0)$$
$$SE(\hat{T} \mid H_0) = \sqrt{\sum_j W_j^2 Var(Z_j^* \mid H_0)}$$

To compute the balancing critical value(C.4), we also need $E(\hat{T}|H_a)$ and $SE(\hat{T}|H_a)$. These values are determined by

$$E(\hat{T}|H_a) = \sum_j W_j E(Z_j^* \mid H_a), \text{ and}$$

$$SE(\hat{T}|H_a) = \sqrt{\sum_j W_j^2 \ var(Z_j^* \mid H_a)}.$$

In which case equation (C.4) gives

$$c_B = \frac{\sum_j W_j E(Z_j^* \mid H_a) - \sum_j W_j E(Z_j^* \mid H_0)}{\sqrt{\sum_j W_j^2 \ var(Z_j^* \mid H_a)} + \sqrt{\sum_j W_j^2 \ var(Z_j^* \mid H_0)}}. \tag{C.5}$$

Thus, we need to determine how to calculate $E(Z_j^* \mid H_0)$, $Var(Z_j^* \mid H_0)$, $E(Z_j^* \mid H_a)$, and $Var(Z_j^* \mid H_a)$. These values depend on the distribution of Z_j (see Appendix A) under the null and alternative hypotheses.

One possible set of hypotheses, that take into account the assumption that transaction are identically distributed within cells, is:

H_0: $\mu_{1j} = \mu_{2j}$, $\sigma_{1j}^2 = \sigma_{2j}^2$

H_a: $\mu_{2j} = \mu_{1j} + \delta_j \cdot \sigma_{1j}$, $\sigma_{2j}^2 = \lambda_j \cdot \sigma_{1j}^2$ $\qquad$ $\delta_j > 0$, $\lambda_j \geq 1$ and $j = 1,\ldots,L$.

Under this null hypothesis, Z_j has a standard normal distribution within each cell j. In which case,

$$E(Z_j^* \mid H_0) = -\frac{1}{\sqrt{2\pi}}, \text{ and}$$
$$var(Z_j^* \mid H_0) = \frac{1}{2} - \frac{1}{2\pi}.$$

Under the alternative hypothesis, Z_j has a normal distribution with

$$E(Z_j \mid H_a) = m_j = \frac{-\delta_j}{\sqrt{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}}, \text{ and}$$

$$SE(Z_j \mid H_a) = se_j = \sqrt{\frac{\lambda_j n_{1j} + n_{2j}}{n_{1j} + n_{2j}}}$$

In general, the mean of a normal distribution truncated at 0 is

$$M(\mu,\sigma) = \int_{-\infty}^{0} \frac{x}{\sqrt{2\pi}\sigma} \exp\left(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2\right) dx,$$

and the variance is

$$V(\mu,\sigma) = \int_{-\infty}^{0} \frac{x^2}{\sqrt{2\pi}\sigma} \exp\left(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2\right) dx - M(\mu,\sigma)^2$$

It can be shown that

$$M(\mu,\sigma) = \mu\Phi\left(\tfrac{-\mu}{\sigma}\right) - \sigma\phi\left(\tfrac{-\mu}{\sigma}\right)$$

and

$$V(\mu,\sigma) = (\mu^2 + \sigma^2)\Phi\left(\tfrac{-\mu}{\sigma}\right) - \mu\sigma\phi\left(\tfrac{-\mu}{\sigma}\right) - M(\mu,\sigma)^2$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function, and $\phi(\cdot)$ is the standard normal density function.

Using the above notation, and equation (C.5), we get the formula for the balancing critical of Z^T for the alternative hypothesis defined above.

$$c_B = \frac{\sum_j W_j M(m_j, se_j) - \sum_j W_j \frac{-1}{\sqrt{2\pi}}}{\sqrt{\sum_j W_j^2 V(m_j, se_j)} + \sqrt{\sum_j W_j^2 \left(\frac{1}{2} - \frac{1}{2\pi}\right)}}. \qquad \text{(C.6)}$$

This formula assumes that Z_j, is approximately normally distributed within cell j. When the cell sample sizes, n_{1j} and n_{2j}, are small this may not be true. It is possible to determine the cell mean and variance under the null hypothesis when the cell sample sizes are small. It is much more difficult to determine these values under the alternative hypothesis. Since the cell weight, W_j will also be small (see Appendix A) for a cell with small volume, the cell mean and variance will not contribute much to the weighted sum. Therefore, formula (C.6) should provide a reasonable approximation to the balancing critical value.

Determining the Parameters of the Alternative Hypothesis

In this appendix we have indexed the alternative hypothesis by two sets of parameters, λ_j and δ_j. While statistical science can be used to evaluate the impact of different choices of these parameters, there is not much that an appeal to statistical principles can offer in directing specific choices. Specific choices are best left to telephony experts. Still, it is possible to comment on some aspects of these choices:

- Parameter Choices for λ_j. The set of parameters λ_j index alternatives to the null hypothesis that arise because there might be greater unpredictability or variability in the delivery of service to a CLEC customer over that which would be achieved for an otherwise comparable ILEC customer. While concerns about differences in the variability of service are important, it turns out that the truncated Z testing which is being recommended here is relatively insensitive to all but very large values of the λ_j. Put another way, reasonable differences in the values chosen here could make very little difference in the balancing points chosen.

- Parameter Choices for δ_j. The set of parameters δ_j are much more important in the choice of the balancing point than was true for the λ_j. The reason for this is that they directly index differences in average service. The truncated Z test is very sensitive to any such differences; hence, even small disagreements among experts in the choice of the δ_j could be very important. Sample size matters here too. For example, setting all the δ_j to a single value — $\delta_j = \delta$ — might be fine for tests across individual CLECs where currently in Louisiana the CLEC customer bases are not too different. Using the same value of δ for the overall state testing does not seem sensible, however, since the state sample would be so much larger.

The bottom line here is that beyond a few general considerations, like those given above, a principled approach to the choice of the alternative hypotheses to guard against must come from elsewhere.

Attachment 2

Balancing the Type I and Type II Error Probabilities of the Modified Z Test Statistic

This paper describes the methodology for balancing the error probabilities when the Modified Z statistic is used for performance measure parity testing. There are four key elements of the statistical testing process:

1. the null hypothesis, H_0, that parity exists between ILEC and CLEC services,
2. the alternative hypothesis, H_a, that the ILEC is giving better service to its own customers,
3. the Modified Z test statistic, Z, and
4. a critical value, c.

The decision rule[6] is

- If $Z < c$, then accept H_a.
- If $Z \geq c$, then accept H_0.

There are two types of error possible when using such a decision rule:

Type I Error: Deciding favoritism exists (accept H_a) when there is, in fact, no favoritism (H_0 is true).
Type II Error: Deciding parity exists (accept H_0) when there is, in fact, favoritism (H_a is true).

The probabilities of the two types of error are:

Type I Error: $\alpha = \mathrm{P}(Z < c \mid H_0)$.
Type II Error: $\beta = \mathrm{P}(Z \geq c \mid H_a)$.

In what follows, we show how to find a balancing critical value, c_B, so that $\alpha = \beta$.

General Methodology

The general form of the test statistic that is being used is

$$z_0 = \frac{\hat{T} - E(\hat{T} \mid H_0)}{SE(\hat{T} \mid H_0)}, \tag{1}$$

[6] This decision rule assumes that the smaller a performance measure is, the better the service. If the opposite is true, then the decision rule should be reversed by using $-Z$ in place of Z.

where

$\hat{T}$ is an estimator that is (approximately) normally distributed,

$E(\hat{T} \mid H_0)$ is the expected value (mean) of $\hat{T}$ under the null hypothesis, and

$SE(\hat{T} \mid H_0)$ is the standard error of $\hat{T}$ under the null hypothesis.

Thus, under the null hypothesis, z_0 follows a standard normal distribution. However, this is not true under the alternative hypothesis. In this case,

$$z_a = \frac{\hat{T} - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)}$$

has (approximately) a standard normal distribution. Here

$E(\hat{T} \mid H_a)$ is the expected value (mean) of $\hat{T}$ under the alternative hypothesis, and

$SE(\hat{T} \mid H_a)$ is the standard error of $\hat{T}$ under the alternative hypothesis.

Notice that

$$\begin{aligned} \beta &= \mathrm{P}(z_0 > c \mid H_a) \\ &= \mathrm{P}\left(z_a > \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)} \right), \end{aligned} \qquad (2)$$

and recall that for a standard normal random variable z and a constant b, $\mathrm{P}(z < b) = \mathrm{P}(z > -b)$. Thus,

$$\alpha = \mathrm{P}(z_0 < c) = \mathrm{P}(z_0 > -c). \qquad (3)$$

Since we want $\alpha = \beta$, the right hand sides of (2) and (3) represent the same area under the standard normal density. Therefore, it must be the case that

$$-c = \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)}.$$

Solving this for c give the general formula for a balancing critical value:

$$c_B = \frac{E(\hat{T} | H_a) - E(\hat{T} | H_0)}{SE(\hat{T} | H_a) + SE(\hat{T} | H_0)}. \tag{4}$$

The Balancing Critical Value of the Modified Z for a Mean Measure

The modified Z statistic, Z, for a mean measure is given by

$$Z = \frac{\hat{T}}{s_1 \sqrt{1/n_1 + 1/n_2}},$$

where $\hat{T} = \overline{X}_1 - \overline{X}_2$, and subscripts 1 and 2 refer to ILEC and CLEC quanti ties, respectively.

One possible set of hypotheses that take into account the assumption that transaction are identically distributed within LECs, is:

H_0: $\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$,

H_a: $\mu_2 = \mu_1 + \delta \cdot \sigma_1$, $\sigma_2^2 = \lambda \cdot \sigma_1^2$, where $\delta > 0$ and $\lambda \geq 1$.

Assuming that n_1 is large enough so that s_1 adequately approximates σ_1, we have

$$E(\hat{T} | H_0) = 0,$$

$$SE(\hat{T} | H_0) = \sigma_1 \sqrt{1/n_1 + 1/n_2},$$

$$E(\hat{T} | H_a) = -\delta \quad _1,$$

$$SE(\hat{T} | H_a) = \sigma_1 \sqrt{1/n_1 + \lambda/n_2}.$$

Substituting these values in equation (4) gives

$$c_B = \frac{-\delta}{\sqrt{1/n_1 + 1/n_2} + \sqrt{1/n_1 + \lambda/n_2}}$$

$$= \frac{-\delta\sqrt{n_1 n_2}}{\sqrt{n_1 + n_2} + \sqrt{\lambda n_1 + n_2}}.$$

The preceding equations have indexed the alternative hypothesis by two parameters, λ and δ. While statistical science can be used to evaluate the impact of different choices of these parameters, there is not much that an appeal to statistical principles can offer in directing specific choices. Specific choices are best left to telephony experts. Still, it is possible to comment on some aspects of these choices:

Parameter Choice for λ. The parameter λ indexes an alternative to the null hypothesis that arises because there might be greater unpredictability or variability in the delivery of service to a CLEC customer over that which would be achieved for an otherwise comparable ILEC customer. Typically, there is little basis for choosing a value of λ other than 1, in which case the formula for c_B simplifies to

$$c_B = \frac{-\delta\sqrt{n_1 n_2}}{2\sqrt{n_1 + n_2}}.$$

Parameter Choice for δ. The parameter δ is much more important in the choice of the balancing point than was true for λ because it directly indexes the difference in average service.

The Balancing Critical Value of the Modified Z for a Proportion Measure

Specification of a balancing critical value for a proportion measure is more complex than for mean measures because c_B depends directly on both the assumed ILEC and CLEC proportions under H_a not just through a single parameter like δ.

The modified Z statistic for a proportion measure is given by

$$Z = \frac{\hat{T}}{\sqrt{\hat{p}_{ILEC}(1-\hat{p}_{ILEC})}\sqrt{1/n_1 + 1/n_2}},$$

where $\hat{T} = \hat{p}_{ILEC} - \hat{p}_{CLEC}$, and where n_1 and n_2 are the ILEC and CLEC sample sizes, respectively.

The null and alternative hypotheses are specified fully in terms of the true proportions p_{ILEC} and p_{CLEC} as follows:

$$H_0\colon\ p_{ILEC} = p_{CLEC} = p_1\ ,$$

$$H_a\colon\ p_{ILEC} = p_1,\ p_{CLEC} = p_2 > p_1.$$

Assuming that n_1 is large enough so that $\hat{p}_{ILEC}(1-\hat{p}_{ILEC})$ adequately approximates $p_{ILEC}(1-p_{ILEC})$, then Z satisfies (1) and we have

$$E(\hat{T} \mid H_0) = 0,$$

$$SE(\hat{T} \mid H_0) = \sqrt{p_1(1-p_1)}\sqrt{1/n_1 + 1/n_2}\,,$$

$$E(\hat{T} \mid H_a) = p_1 - p_2\,,$$

$$SE(\hat{T} \mid H_a) = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}\,.$$

Substituting these values in equation (4) gives

$$c_B = \frac{-(p_2 - p_1)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} + \sqrt{p_1(1-p_1)}\sqrt{1/n_1 + 1/n_2}}\,.$$

A convenient way to specify the alternative hypothesis is through the "odds ratio" for p_2 and p_1; specifically

$$\varphi = \left(\frac{p_2}{p_1}\right)\left(\frac{1-p_1}{1-p_2}\right),$$

so that

$$p_2 = \left(\frac{\varphi p_1}{1 + (\varphi - 1)p_1}\right).$$

The Balancing Critical Value of the Modified Z for a Rate Measure

A rate is a ratio of two counts *num/denom*—e.g., $r_{ILEC} = num_{ILEC}/denom_{ILEC}$. Where the *denom* count is assumed known but the *num* count is subject to sampling variability. Similarly to proportions, the balancing critical value c_B depends directly on the assumed ILEC and CLEC rates under H_a as well as the ILEC and CLEC denominators.

The modified Z statistic for a rate measure is given by

$$Z = \frac{\hat{T}}{\sqrt{\hat{r}_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})}}\,.$$

Where $\hat{T} = \hat{r}_{ILEC} - \hat{r}_{CLEC}$.

The null and alternative hypotheses are specified fully in terms of the true proportions r_{ILEC} and r_{CLEC} as follows:

H_0: $r_{ILEC} = r_{CLEC} = r_1$,

H_a: $r_{ILEC} = r_1$, $r_{CLEC} = r_2 > r_1$.

Assuming that $denom_{ILEC}$ is large enough so that $\hat{r}_{ILEC}$ adequately approximates r_{ILEC}, then Z satisfies (1) and we have

$$E(\hat{T} \mid H_0) = 0,$$

$$SE(\hat{T} \mid H_0) = \sqrt{r_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})},$$

$$E(\hat{T} \mid H_a) = r_1 - r_2,$$

$$SE(\hat{T} \mid H_a) = \sqrt{r_{CLEC}/denom_{CLEC} + r_{ILEC}/denom_{ILEC})}.$$

Substituting these values in equation (4) gives

$$c_B = \frac{-(r_2 - r_1)}{\sqrt{r_{CLEC}/denom_{CLEC} + r_{ILEC}/denom_{ILEC})} + \sqrt{r_{ILEC}(1/denom_{CLEC} + 1/denom_{ILEC})}}.$$

A convenient way to specify the alternative hypothesis is by

$$r_2 = \varepsilon r_1 .$$

Exhibit B

Local Competition Users Group

Statistical Tests for Local Service Parity

February 6, 1998
Membership: AT&T, Sprint, MCI, LCI, WorldCom

Version 1.0

Executive Summary

The Local Competition Users Group has drafted 27 Service Quality Measurements (SQMs) that will be used to measure parity of service provided by incumbent local exchange carriers (ILECs) to competitive local exchange carriers (CLECs). This set of measures includes means, proportions, and rates of various indicators of service quality. This document proposes statistical tests that are appropriate for determining if parity is being provided with respect to these measurements.

Each month, a specified report of the 27 SQMs will be provided by the ILEC, broken down by the requested reporting dimensions. The SQMs are to be systematically developed and provided by the ILECs as specified. Test parameters will be calculated so that the overall probability of declaring the ILEC to be out of parity purely by chance is very small. For each SQM and reporting dimension reported, the difference between the ILEC and CLEC results is converted to a *z*-value. Non-parity is determined if a *z*-value exceeds a selected critical value.

Introduction

Purpose

The Local Competition Users Group (LCUG) is a cooperative effort of AT&T, MCI, Sprint, LCI and WorldCom for establishing standards for the entry of new companies (competitive local exchange carriers, or CLECs) into the local telecommunications market. A key initiative of the LCUG is to establish measures of parity for services provided by incumbent local exchange carriers (ILECs). In short, parity means that the support ILECs provide on behalf of the CLECs is no lesser in quality than the service provided by the ILECs to their own customers.

The LCUG has drafted a document listing service quality measurements (SQMs) that must be reported by the ILECs to insure that CLECs are given parity of support. The SQM document has been submitted to the FCC and made available to PUCs in all 50 states and is pending approval by many of these regulatory agencies. This document has been drafted to describe statistical methodology for determining if parity exists based on the measurements defined in the SQM document.

Service Quality Measurements

The LCUG has identified 27 service quality measurements for testing parity of service. These are:

Category	ID	Description
Pre-Ordering	PO-1	Average Response Interval for Pre-Ordering Information
Ordering and Provisioning	OP-1	Average Completion Interval
	OP-2	Percent Orders Completed on Time
	OP-3	Percent Order Accuracy
	OP-4	Mean Reject Interval
	OP-5	Mean FOC Interval
	OP-6	Mean Jeopardy Interval
	OP-7	Mean Completion Interval
	OP-8	Percent Jeopardies Returned
	OP-9	Mean Held Order Interval
	OP-10	Percent Orders Held >= 90 Days
	OP-11	Percent Orders Held >= 15 Days
Maintenance and Repair	MR-1	Mean Time to Restore
	MR-2	Repeat Trouble Rate
	MR-3	Trouble Rate
	MR-4	Percentage of Customer Troubles Resolved

		Within Estimate
General	GE-1	Percent System Availability
	GE-2	Mean Time to Answer Calls
	GE-3	Call Abandonment Rate
Billing	BI-1	Mean Time to Provide Recorded Usage Records
	BI-2	Mean Time to Deliver Invoices
	BI-3	Percent Invoice Accuracy
	BI-4	Percent Usage Accuracy
Operator Services and Directory Assistance	OSDA-1	Mean Time to Answer
Network Performance	NP-1	Network Performance Parity
Interconnect / Unbundled Elements and Combos	IUE-1	Function Availability
	IUE-2	Timeliness of Element Performance

The Service Quality Measurements document describes the importance of each measure as an indicator of service parity. The SQM document also describes reporting dimensions that will be used to break each measure out by like factors (*e.g.*, major service group).

Why We Need to Use Statistical Tests

The Telecommunications Act of 1996 requires that ILECs provide nondiscriminatory support regardless of whether the CLEC elects to employ interconnection, services resale, or unbundled network elements as the market entry method. It is essential that CLECs and regulators be able to determine whether ILECs are meeting these parity and nondiscriminatory obligations. In order to make such a determination, the ILEC's performance for itself must be compared to the ILEC's performance in support of CLEC operations; and the results of this comparison must demonstrate that the CLEC receives no less than equal treatment compared to that the ILEC provides to its own operations. Where a direct comparison to analogous ILEC performance is not possible, the comparative standard is the level of performance that offers an efficient CLEC a meaningful opportunity to compete.

When making the comparison of ILEC results to CLEC results, it is necessary to employ comparative procedures that are based upon generally accepted statistical procedures. It is important to use statistical procedures because all of the ILEC-CLEC processes that will be measured are processes that contain some degree of randomness. Statistical procedures recognize that there is measurement variability, and assist in translating results data into useful decision-making information. A statistical approach allows for measurement variability while controlling the risk of drawing an inappropriate conclusion (*i.e,* a "type 1" or "type 2" error, discussed in the next section).

Basic Concepts and Terms

Populations and Samples

Statistical procedures will permit a determination whether the support that the ILECs provide to CLECs is indistinguishable from the support provided by the ILECs to their own customers. In statistical terms, we will determine whether two "samples", the ILEC sample and the CLEC sample, come from the same "population" of measurements.

The procedures described in this paper are based on the following assumption: *When parity is provided, the ILEC data and CLEC data can both be regarded as samples from a common population of possible outcomes.* In other words, if parity exists, the measured results for a CLEC should not be distinguishable from the measured results for the ILEC, once random variability is taken into account. Figure 1 illustrates this concept. On the right side of the figure are histograms of two samples. In this illustration, the ILEC sample contains 200 observations (data values) and the CLEC sample contains 50. Note that the two histograms are not exactly alike. This is due to sampling variation. The assumption that parity exists implies that both samples were drawn from the same population of values. If it were possible to observe this population completely, the population histogram might appear as shown on the left of the Figure. If the samples were indeed taken from this population, histograms drawn for larger and larger samples would look more and more like the population histogram. Figure 1 shows that even when parity is being provided, there will be differences between the samples due to sampling variability. Statistical tests quantify the differences between the two samples and make proper allowance for sampling variability. They assess the chance that the differences that are observed are due simply to sampling variability, if parity is being provided.

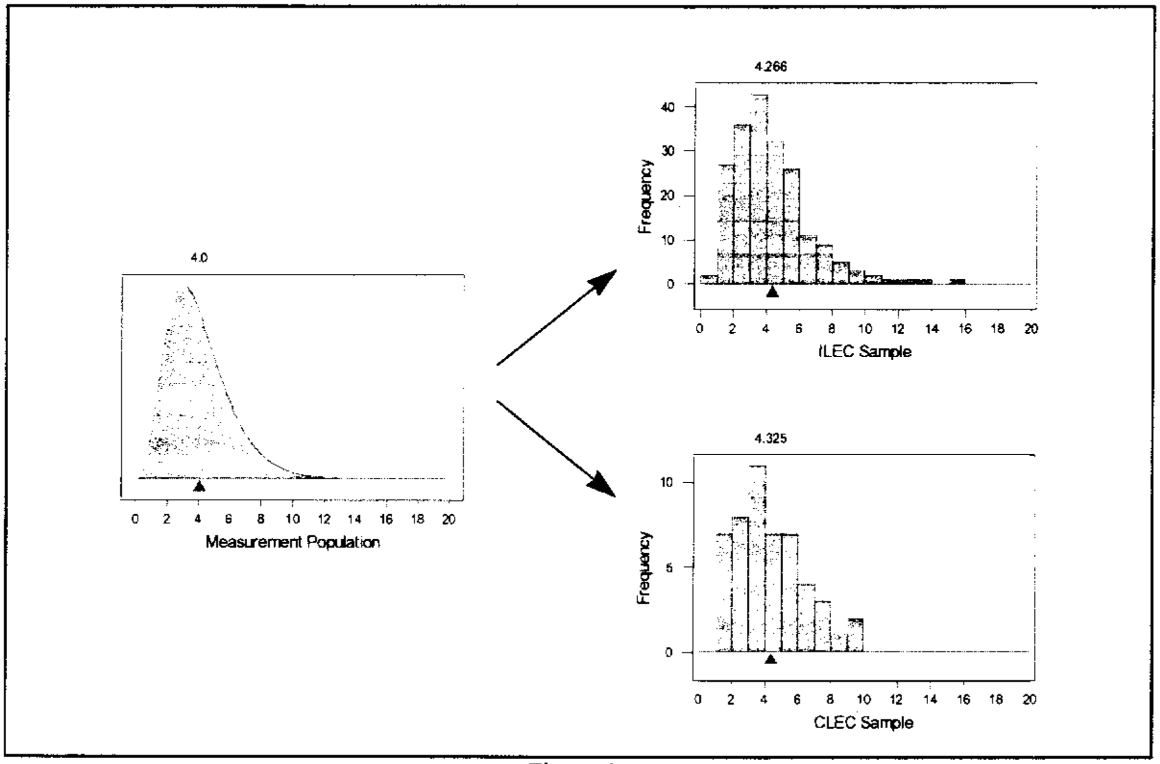

Figure 1.

Measures of Central Tendency and Spread

Often, distributions are summarized using "statistics." For the purpose of this paper, a "statistic" is simply a calculation performed on a sample set of data. Two common types of statistics are known as measures of "central tendency" and "spread."

A measure of central tendency is a summary calculation that describes the middle of the distribution in some way. The most common measure of central tendency is called the "mean" or "average" of the distribution. The mean of a sample is simply the sum of the data values divided by the sample size (number of observations). Algebraically, this calculation is expressed as

$$\bar{x} = \frac{\Sigma x}{n},$$

where x denotes a value in the sample and n denotes the sample size. The mean describes the center of the distribution in the following way: *If the histogram for a sample were a set of weights stacked on top of a flat board placed on top of a fulcrum (a "see-saw"), the mean would be the position along the board at which the board would balance.* (See Figure 1.) The mean in Figure 1 is indicated by the small triangle at approximately the value "4" on the horizontal axis.

A measure of spread is a summary calculation that describes the amount of variation in a sample. A common measure of spread is a called the "standard deviation" of the sample. The standard deviation is the typical size of a deviation of the observations in the sample from their mean value. The standard deviation is calculated by subtracting the mean value from each observation in the sample, squaring the resulting differences (so that negative and positive differences don't offset), summing the squared differences, dividing the sum by one less than the sample size, then taking the square root of the result. Algebraically, this calculation is expressed as

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}.$$

While the notion of mean and standard deviation exists for populations as well as samples, the mathematical definition for the mean and standard deviation for populations is beyond the scope of this paper. However, their interpretation is generally the same as for samples. In fact, for very large samples, the sample mean and sample standard deviation will be very close to the mean and standard deviation of the population from which the sample was taken.

Sampling Distribution of the Sample Mean

In Figure 1 we showed the positions of the means of the population and the two samples with triangular symbols beneath the distributions. If we sample over successive months, we will get new ILEC samples and new CLEC samples each and every month. These samples will not be exactly like the one for the first month; each will be influenced by sampling variability in a different way. In Figure 2, we show how sets of 100 successive ILEC means and 100 successive CLEC means might appear. The ILEC means can be thought of as being drawn from a population of sample means; this population is called the "sampling distribution" of these ILEC means. This sampling distribution is completely determined by the basic population of measurements that we start with, and the number of observations in each sample. The sampling distribution has the same mean as the population.

Figure 2 illustrates two important statistical concepts:

1. The histogram of successive sample means resembles a bell-shaped curve known as the Normal Distribution. This is true even though the individual observations came from a skewed distribution.
2. The standard deviation of the distribution of sample means is much smaller than the standard deviation of the observations themselves. In fact, statistical theory establishes the fact that the standard deviation on the population of means is smaller by a factor $\sqrt{n}$, where n is the sample size. This effect can be seen in our example: the distribution of the CLEC means is twice as broad

as the distribution of the ILEC means, since the ILEC sample size (200) is four times as large as the CLEC sample size (50).

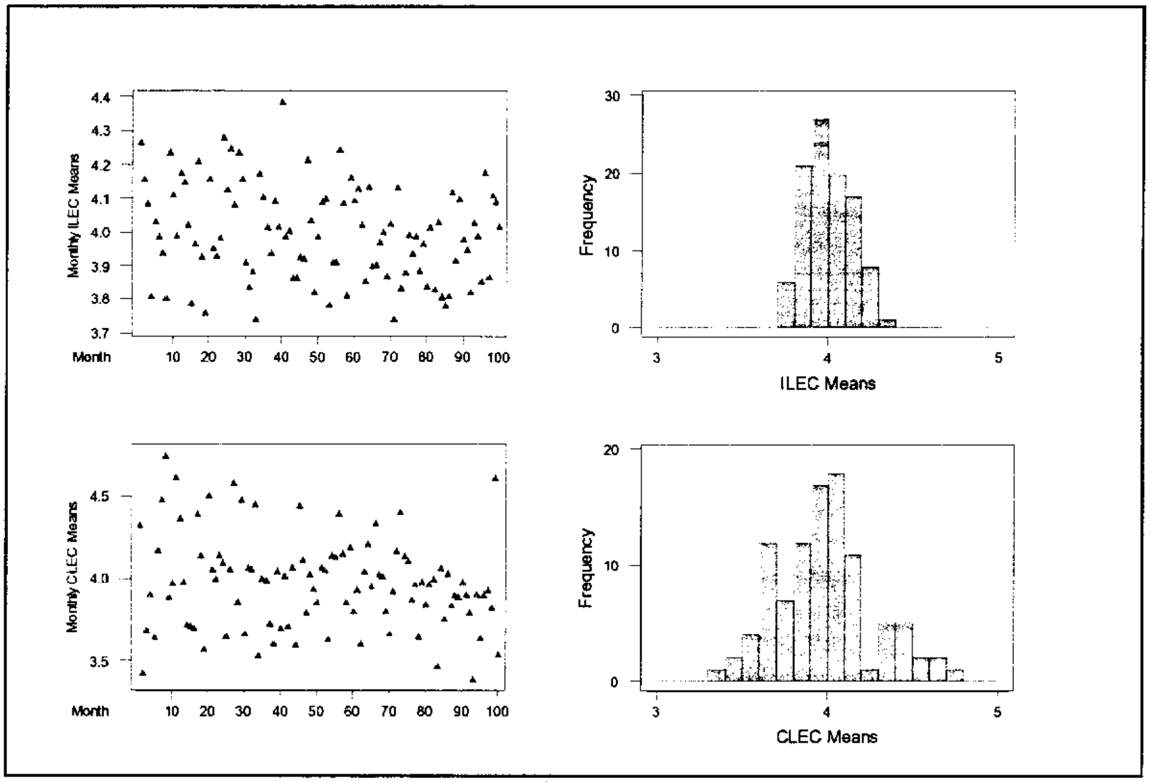

Figure 2.

It is common to call the standard deviation of the sampling distribution of a statistic the "standard error" for the statistic. We shall adopt this convention to avoid confusion between the standard deviation of the individual observations and the standard deviation (standard error) of the statistic. The latter is generally much smaller than the former. In the case of sample means, the standard error of the mean is smaller than the standard deviation of the individual observations by a factor of $\sqrt{n}$.

The Z-test

Our objective is to compare the mean of a sample of ILEC measurements with the mean of a sample of CLEC measurements. Suppose both samples were drawn from the same population; then the difference between these two sample means (*i.e.*, $DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$) will have a sampling distribution which will

(i) have a mean of zero; and
(ii) have a standard error that depends on the population standard deviation and the sizes of the two samples.

Statisticians utilize an index for comparing measurement results for different samples. The index employed is a ratio of the difference in the two sample means (being compared) and the standard deviation estimated for the overall population. This ratio is known as a z-score. The z-score compares the two samples on a standard scale, making proper allowance for the sample sizes.

The computation of the difference in the two sample means is straightforward.

$$DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$$

The standard deviation is less intuitive. Nevertheless, statistical theory establishes the fact that

$$\sigma^2_{\text{DIFF}} = \frac{\sigma^2}{n_{\text{CLEC}}} + \frac{\sigma^2}{n_{\text{ILEC}}},$$

where ? is the standard deviation of the population from which both samples are drawn. That is, the squared standard error of the difference is the sum of the squared standard errors of the two means being compared.[1]

We do not know the true value of the population ? ?because the population cannot be fully observed. However, we can estimate ? given the standard deviation of the ILEC sample (?$_{\text{ILEC}}$).[2] Hence, we may estimate the standard error of the difference with

$$\sigma_{\text{DIFF}} = \sqrt{\frac{\sigma^2_{\text{ILEC}}}{n_{\text{CLEC}}} + \frac{\sigma^2_{\text{ILEC}}}{n_{\text{ILEC}}}} = \sqrt{\sigma^2_{\text{ILEC}}\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

If we then divide the difference between the two sample means by this estimate of the standard deviation of this difference, we get what is called a "z-score".

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

Because we assumed that both samples were in fact drawn from the same population, this *z*-score has a sampling distribution that is very nearly Standard Normal, *i.e.*, having a mean of zero and a standard error of one. Thus, the *z*-score will lie between ± 1 in about 68% of cases, will lie between ± 2 in about 95% of cases, and will lie between ± 3 in about 99.7% of cases, always

[1] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 370.

[2] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 338.

assuming that both samples come from the same population. Therefore, one possible procedure for checking whether both samples come from the same population is to compare the *z*-score with some cut-off value, perhaps +3. For comparisons where the values of z exceed the cutoff value, you reject the assumption of parity as not proven by the measured results. This is an example of a statistical test procedure. It is a formal rule of procedure, where we start with raw data (here two samples, ILEC measurements and CLEC measurements), and arrive at a decision, either "conformity" or" violation ".

Type 1 Errors and Type 2 Errors

Each statistical test has two important properties. The first is the probability that the test will determine that a problem exists when in fact there is none. Such a mistaken conclusion is called a type one error. In the case of testing for parity, a type one error is the mistake of charging the ILEC with a parity violation when they may not be acting in a discriminatory manner. The second property is the probability that the test procedure will not identify a parit y violation when one does exist. The mistake of not identifying parity violation when the ILEC is providing discriminatory service is called a type two error. A balanced test is, therefore, required.

From the ILEC perspective, the statistical test proce dure will be unacceptable if it has a high probability of type one errors. From the CLEC perspective, the test procedure will be unacceptable if it has a high probability of type two errors.

Very many test procedures are available, all having the same pr obability of type one error. However the probability of a type two error depends on the particular kind of violation that occurs. For small departures from parity, the probability of detecting the violation will be small. However, different test procedu res will have different type two error probabilities. Some test procedures will have small type two error when the CLEC mean is larger than the ILEC mean, even if the CLEC standard deviation is the same as the ILEC standard deviation, while other procedures will be sensitive to differences in standard deviation, even if the means are equal. Our proposals below are designed to have small type two error when the CLEC mean exceeds the ILEC mean, whether or not the two variances are equal.

Tests of Proportions and Rates

When our measurements are proportions (*e.g.* percent orders completed on time) rather than measurements on a scale, there are some simplifications. We can think of the "population" as being analogous to an urn filled with balls, each labeled either 0(failure) or 1(success). In this population, the fraction of 1's is some "population proportion". Making an observation corresponds to drawing a single ball from this urn. Each month, the ILEC makes some number of observations, and reports th e ratio of failures or successes to the total number of observations; the ILEC does the same does the same for the CLEC. The situation is very similar to that discussed above; however, rather than a wide range of possible result values, we simply have 0s (failures) and 1s (successes). The "sample mean" becomes the "observed proportion", and this will have a sampling distribution just as before. The novelty of the situation is that now the population standard deviation is a known function of the populat ion proportion[3]; if the population proportion is p, the population standard deviation is $\sqrt{p(1-p)}$, with similar simplifications in all the other formulas.

There is a similar simplification when the observations are of rates, *e.g.*, number of troubles per 100 lines. The formulas appear below.

Proposed Test Procedures

Applying the Appropriate Test

Three z-tests will be described in this section: the "Test for Parity in Means", the "Test for Parity in Rates", and the "Test for Parity in Proport ions". For each LCUG Service Quality Measurement (SQM), one or more of these parity tests will apply. The following chart is a guide that matches each SQM with the appropriate test.

Measurement (Corresponding LCUG Number)	*Test*
Preordering Response Interval (PO-1)	**Mean**
Avg. Order Completion Interval (OP-1)	**Mean**
% Orders Completed On Time (OP-2)	**Proportion**
% Order (Provisioning) Accuracy (OP-3)	**Proportion**
Order Reject Interval (OP-4)	**Mean**
Firm Order Confirmation Interval (OP -5)	**Mean**
Mean Jeopardy Interval (OP-6)	**Mean**
Completion Notice Interval (OP-7)	**Mean**
Percent Jeopardies Returned (OP-8)	**Proportion**
Held Order Interval (OP-9)	**Mean**
% Orders Held ≥ 90 Days (OP-10)	**Proportion**
% Orders Held ≥ 15 Days (OP-11)	**Proportion**
Time To Restore (MR-1)	**Mean**
Repeat Trouble Rate (MR-2)	**Proportion**
Frequency of Troubles (MR-3)	**Rate**
Estimated Time To Restore (MR-4)	**Proportion**
System Availability (GE-1)	**Proportion**
Center Speed of Answer (GE-2)	**Mean**
Call Abandonment Rate (GE-3)	**Proportion**
Mean Time to Deliver Usage Records (BI-1)	**Mean**
Mean Time to Deliver Invoices (BI-2)	**Mean**
Percent Invoice Accuracy (BI-3)	**Proportion**
Percent Usage Accuracy (BI-4)	**Proportion**

[3] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 212.

OS/DA Speed of Answer (OS/DA-1)	**Mean**
Network Performance (NP-1)	**Mean, Proportion**
Availability of Network Elements (IUE-1)	**Mean, Proportion**
Performance of Network Elements (IUE-2)	**Mean, Proportion**

Test for Parity in Means

Several of the measurements in the LCUG SQM document are averages (*i.e.*, means) of certain process results. The statistical procedure for testing for parity in ILEC and CLEC means is described below:

1. Calculate for each sample the number of measurements (n_{ILEC} and n_{CLEC}), the sample means ($\bar{x}_{\mathrm{ILEC}}$ and $\bar{x}_{\mathrm{CLEC}}$), and the sample standard deviations (? ILEC and ? CLEC).

2. Calculate the difference between the two sample means; if *larger* CLEC mean indicates possible violation of parity, use $DIFF = \bar{x}_{\mathrm{CLEC}} - \bar{x}_{\mathrm{ILEC}}$, otherwise reverse the order of the CLEC mean and the ILEC mean.

3. To determine a suitable scale on which to measure this difference, we use an estimate of the population variance based on the ILEC sample, adjusted for the sized of the two samples: this gives the standard error of the difference between the means as

$$\sigma_{\mathrm{DIFF}} = \sqrt{\sigma^2_{\mathrm{ILEC}}\left[\frac{1}{n_{\mathrm{CLEC}}} + \frac{1}{n_{\mathrm{ILEC}}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{\mathrm{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
n	mean	variance	n	mean	variance	z	Violation
250	4.038	1.9547	50	5.154	23.2035	5.15	YES!

Test for Parity in Proportions

Several of the measurements in the LCUG SQM document are proportions derived from certain counts. The statistical procedure for testing for parity in ILEC and CLEC proportions is described below. It is the same as that for means, except that we do not need to estimate the ILEC variance separately.

1. Calculate for each sample sample sizes (n_{ILEC} and n_{CLEC}), and the sample proportions (p_{ILEC} and p_{CLEC}).

2. Calculate the difference between the two sample means; if *larger* CLEC proportion indicates worse performance, use $DIFF = p_{\text{CLEC}} - p_{\text{ILEC}}$, otherwise reverse the order of the ILEC and CLEC proportions.

3. Calculate an estimate of the *standard error for the difference* in the two proportions according to the formula

$$\sigma_{\text{DIFF}} = \sqrt{p_{\text{ILEC}}(1 - p_{\text{ILEC}})\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

4. Hence compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
num	den	p	num	den	p	z	Violation
5	250	2.00%	7	40	17.50%	6.50	YES!

Test for Parity in Rates

A rate is a ratio of two counts, *num/denom*. An example of this is the trouble rate experience for POTS. The procedure for analyzing measurements results that are rates is very similar to that for proportions.

1. Calculate the numerator and the denominator counts for both ILEC and CLEC, and hence the two rates $r_{\text{ILEC}} = num_{\text{ILEC}}/denom_{\text{ILEC}}$ and $r_{\text{CLEC}} = num_{\text{CLEC}}/denom_{\text{CLEC}}$.

2. Calculate the difference between the two sample rates; if *larger* CLEC rate indicates worse performance, use $DIFF = r_{\text{CLEC}} - r_{\text{ILEC}}$, otherwise take the negative of this.

3. Calculate an estimate of the *standard error for the difference* in the two rates according to the formula

$$\sigma_{\text{DIFF}} = \sqrt{r_{\text{ILEC}}\left[\frac{1}{denom_{\text{CLEC}}} + \frac{1}{denom_{\text{ILEC}}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
num	den	rate	num	den	rate	z	Violation
250	610	0.409836	34	30	1.133333	6.04	YES!